

We Claim:

1. An arrangement for creating synthesized speech from an applied sequence of desired speech unit features parameter sets, D-SUF(i), $i=2,3,\dots$, comprising:

a database that contains a plurality of sets, $E(k)$, $k=1,2,\dots,K$, where K is an integer, each set including

a collection of unit selection parameters that characterize a unit of speech of set $E(k)$,

and a plurality of frames associated with said $E(k)$, each of said frames being represented by

a collection of model parameters, and

T-D data representing a time-domain speech signal corresponding to said frame

a database search engine that, for each applied D-SUF(i), selects from said database a set $E(i)$ having a collection of unit selection parameters that match best said D-SUF(i), and said plurality of frames that are associated with said $E(i)$, thus creating a sequence of frames;

an evaluator that determines, based on assessment of information obtained from said database and pertaining to said $E(i)$, whether modifications are needed to frames of said $E(i)$;

a modification and synthesis module that, when said evaluator concludes that modifications to frames are needed, modifies the collection of model parameters of those frames that need modification, and generates, for each frame having a modified collection of model parameters, T-D data corresponding to said frame; and

a combiner that concatenates T-D data of successive frames in said sequence of frames, by employing, for each concatenated frame, the T-D data generated for said concatenated frame by said modification and synthesis module, if such T-D data was generated, or T-D data retrieved for said concatenated frame from said database.

2. The arrangement of claim 1 where said assessment by said evaluator is made with a comparison between collection of model parameters of a frame at a head end of

said $E(i)$ and collection of model parameter of a frame at a tail end of a previously selected set, $E(i-1)$.

3. The arrangement of claim 2 where said comparison determines whether said model parameters of said frame at head end of said $E(i)$ differ from said model parameters of said frame at a tail end of said $E(i-1)$ by more than a preselected amount.

4. The arrangement of claim 3 where said comparison is based on fundamental frequency of said frame at head end of said $E(i)$ and fundamental frequency of said frame at a tail end of said $E(i-1)$.

5. The arrangement of claim 2 where said modification and synthesis module modifies, when said evaluator determines that modifications to frames are needed, collections of model parameters of a first chosen number of frames that are at a head region of said $E(i)$, and collections of model parameters of a second chosen number of frames that are at a tail region of said $E(i-1)$.

6. The arrangement of claim 2 where said modification and synthesis unit modifies said collections of model parameters of said first chosen number of frames that are at a head region of said $E(i)$, and collections of model parameters of said second chosen number of frames that are at a tail region of said $E(i-1)$ in accordance with an interpolation algorithm.

7. The arrangement of claim 6 where said interpolation algorithm interpolates fundamental frequency parameter of the modified collections of model parameters.

8. The arrangement of claim 6 where said interpolation algorithm interpolates fundamental frequency parameter and amplitude parameters of the modified collections of model parameters.

9. The arrangement of claim 1 said assessment by said evaluator is made with a comparison between unit selection parameters of E(i) and said D-SUF(i).

10. The arrangement of claim 9 where said comparison determines where said unit selection parameters of said selected set E(i) differ from said D-SUF(i) by more than a selected threshold.

11. The arrangement of claim 9 where said modification and synthesis module modifies, when said evaluator determines that modifications to frames are needed, the collections of model parameters of frames of said E(i).

12. The arrangement of claim 1 where said assessment by said evaluator is made with a first comparison between unit selection parameters of E(i) and said D-SUF(i) and with a second comparison between collection of model parameters of a frame at a head end of said E(i) and collection of model parameter of a frame at a tail end of a previously selected set, E(i-1).

13. The arrangement of claim 12 where in said second comparison, said frame at a head end of said E(i) is considered after taking account of modifications to said collection of model parameters of said frame at the head end of E(i) pursuant to said first comparison.

14. The arrangement of claim 1 where said T-D data stored in said database represents one pitch period of speech, said T-D data generated by said modification and synthesis module represents one pitch period of speech, and said combiner concatenates T-D data of a frame by creating additional data for said frame to form an extended speech representation of associated frames, and carrying out a filtering and an overlap-and-add operations to add the T-D data and the created additional data to previously concatenated data.

15. The arrangement of claim 14 where said created additional data extends speech representation to two pitch periods of speech.

16. The arrangement of claim 1 where said T-D data stored in said database in association with a frame is data that was generated from said collection of model parameters associated with said frame.

17. The arrangement of claim 1 where said model parameters of a frame are in accordance with an Harmonic Plus Noise model of speech.

18. The arrangement of claim 1 where durations of said units are related to sounds of said speech segments rather than being preselected at a uniform duration.

19. The arrangement of claim 1 where said model parameters of a frame are obtained from analysis of overlapping speech frames that are on the order of two pitch periods each for voiced speech.

20. The arrangement of claim 1 further comprising a text-to-speech units converter for developing said D-SUF(i), $i=2,3,\dots$

21. The arrangement of claim 1 where said database search engine, evaluator, modification and synthesis module, and combiner are software modules executing on a stored program processor.

22. A method for creating synthesized speech from an applied sequence of desired speech unit features parameter sets, D-SUF(i), $i=2,3,\dots$, comprising the steps pf:

for each of said D-SUF(i), selecting from a database information of an entry E(i) having a set of speech unit characterization parameters that best match said D-SUF(i), which entry also includes a plurality of frames represented by a corresponding plurality of model parameter sets, and a corresponding plurality of time domain speech frames, said information including at least said plurality of model parameter sets, thereby

resulting in a sequence of model parameter sets, corresponding to which a sequence of output speech frames is to be concatenated;

determining, based on assessment of information obtained from said database and pertaining to said $E(i)$, whether modifications are needed to said frames of said $E(i)$;

when said evaluator concludes that modifications to frames are needed, modifying the collection of model parameters of those frames that need modification,

generating, for each frame having a modified collection of model parameters, T-D data corresponding to said frame; and

concatenating T-D data of successive frames in said sequence of frames, by employing, for each concatenated frame, the T-D data generated for said step of generating, if such T-D data was generated, or T-D data retrieved for said concatenated frame from said database.

23. The method of claim **22** where said assessment by said evaluator is made with a comparison between collection of model parameters of a frame at a head end of said $E(i)$ and collection of model parameter of a frame at a tail end of a previously selected set, $E(i-1)$.

24. The method of claim **23** where said comparison determines whether said model parameters of said frame at head end of said $E(i)$ differ from said model parameters of said frame at a tail end of said $E(i-1)$ by more than a preselected amount.

25. The method of claim **24** where said comparison is based on fundamental frequency of said frame at head end of said $E(i)$ and fundamental frequency of said frame at a tail end of said $E(i-1)$.

26. The method of claim **23** where said modification and synthesis module modifies, when said step of determining concludes that modifications to frames are needed, collections of model parameters of a first chosen number of frames that are at a head region of said $E(i)$, and collections of model parameters of a second chosen number of frames that are at a tail region of said $E(i-1)$.

27. The method of claim **23** where said modification and synthesis unit modifies said collections of model parameters of said first chosen number of frames that are at a head region of said $E(i)$, and collections of model parameters of said second chosen number of frames that are at a tail region of said $E(i-1)$ in accordance with an interpolation algorithm.

28. The method of claim **27** where said interpolation algorithm interpolates fundamental frequency parameter of the modified collections of model parameters.

29. The method of claim **27** where said interpolation algorithm interpolates fundamental frequency parameter and amplitude parameters of the modified collections of model parameters.

30. The method of claim **22** said assessment by said step of determining is made with a comparison between unit selection parameters of $E(i)$ and said $D-SUF(i)$.

31. The method of claim **30** where said comparison determines where said unit selection parameters of said selected set $E(i)$ differ from said $D-SUF(i)$ by more than a selected threshold.

32. The method of claim **30** where said step of modifying modifies, when said determining concludes that modifications to frames are needed, the collections of model parameters of frames of said $E(i)$.

33. The method of claim **22** where said assessment is made with a first comparison between unit selection parameters of $E(i)$ and said $D-SUF(i)$ and with a second comparison between collection of model parameters of a frame at a head end of said $E(i)$ and collection of model parameter of a frame at a tail end of a previously selected set, $E(i-1)$.

34. The method of claim **33** where in said second comparison, said frame at a head end of said E(i) is considered after taking account of modifications to said collection of model parameters of said frame at the head end of E(i) pursuant to said first comparison.

35. The method of claim **22** where said T-D data stored in said database represents one pitch period of speech, said T-D data generated by said step of generating represents one pitch period of speech, and said step of concatenating concatenates T-D data of a frame by creating additional data for said frame to form an extended speech representation of associated frames, and carrying out a filtering and an overlap-and-add operations to add the T-D data and the created additional data to previously concatenated data.

36. The method of claim **35** where said created additional data extends speech representation to two pitch periods of speech.

37. The method of claim **22** where said T-D data stored in said database in association with a frame is data that was generated from said collection of model parameters associated with said frame.

38. The method of claim **22** where said model parameters of a frame are in accordance with an Harmonic Plus Noise model of speech.

39. The method of claim **22** where durations of said units are related to sounds of said speech segments rather than being preselected at a uniform duration.

40. The method of claim **22** where said model parameters of a frame are obtained from analysis of overlapping speech frames that are on the order of two pitch periods each for voiced speech.

41. The method of claim **22** further comprising a step of converting an applied text to a sequence of said D-SUF(i), $i=2,3,\dots$